

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OW protein - nucleic search, using frame\_plus.p2n model

Run on: August 23, 2003, 18:20:39 ; Search time 1902 Seconds  
(without alignments)  
4472.433 Million cell updates/sec

Title: US-09-745-506-37  
Perfect score: 1799  
Sequence: 1 MDKALLSLNDFAISLSPAE.....LENKINIILSETRDPLGVV 350

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=frame+.p2n.model -DEV=xlh  
-Q/cgn2.1/USPRO.spool/US09745506/runat.22082003.104401.7049/app-query.fasta.1.519  
-DB-EST-QFMT=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS-bits -START=1 -END=1 -MATRIX=blomsun62 -TRANS=human40.cgi -LIST=45  
-DOCALLIGN=200 -THR\_SCORE=dot -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MITLEN=0 -MAXLEN=200000000  
-USER=US09745506 -GCCN=1.1.2135 -runat.22082003.104401.7049 -NCP=6 -ICPU=3  
-NO\_MAP -LARGEDEBERG -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGCLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FEAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:\*  
1: em\_estba:\*  
2: em\_estbun:\*  
3: em\_estcin:\*  
4: em\_estcinu:\*  
5: em\_estcov:\*  
6: em\_estcpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estcin:\*  
16: em\_estcinu:\*  
17: em\_estcinu:\*  
18: em\_estcinu:\*  
19: em\_estcinu:\*  
20: em\_estcinu:\*  
21: em\_estcinu:\*  
22: em\_estcinu:\*  
23: em\_estcinu:\*  
24: em\_estcinu:\*  
25: em\_estcinu:\*  
26: em\_estcinu:\*  
27: em\_estcinu:\*  
28: gb\_est1:\*

29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1599	88.9	1201	9	AL529615
2	1548	86.0	1376	11	AK011670
3	1539	85.5	1117	12	BM545164
4	1534	85.3	1201	13	BX446370
5	1481	82.3	1201	9	AL581997
6	1407	78.2	1467	11	AK003978
7	1378	76.6	930	13	BX393871
8	1363	75.8	857	13	BUI72435
9	1362	75.7	929	13	BUI859307
10	1362	75.7	985	13	BQ927763
11	1360.5	75.6	836	13	BQ229243
12	1345.5	74.8	1201	9	AL582089
13	1335	74.2	1201	9	AL563496
14	1316	73.2	1919	11	AK076788
15	1315	73.1	957	9	AL521920
16	1310	72.8	845	10	BG754550
17	1298	72.2	1201	13	BX377118
18	1258	69.9	742	12	BUI093955
19	1251.5	69.6	1201	9	AL523985
20	1233.5	68.6	1019	12	BM557530
21	1233	68.5	890	9	AL520538
22	1201	66.8	888	13	BUI95469
23	1198	66.6	732	10	BG472953
24	1197	66.5	980	13	BQ854847
25	1194	66.4	905	10	BF973755
26	1192.5	66.3	843	12	BI753964
27	1182	65.7	958	10	BF973889
28	1177	65.4	817	12	BI755123
29	1172	65.1	977	10	BF663320
30	1162.5	64.6	1079	13	BX391715
31	1158.5	64.4	963	13	BE797115
32	1157	64.3	759	10	BE753244
33	1135.5	63.1	977	14	CD513847
34	1132.5	63.0	843	13	BQ848893
35	1126.5	62.6	909	14	CD385139
36	1118	62.1	961	9	AL521919
37	1114	61.9	984	13	BX385101
38	1103	61.3	909	13	BUI65031
39	1102	61.3	999	10	BE745082
40	1097	61.0	946	13	BQ963633
41	1082.5	60.2	945	10	BE747311
42	1077	59.9	772	14	CA321994
43	1075	59.8	877	13	BQ437698
44	1069.5	59.4	768	12	BI772801
45	1062	59.0	705	10	BE745354

#### ALIGNMENTS

RESULT 1	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE
AL529615	AL529615	1201 bp mRNA	AL529615	EST	GI:31067458	Homo sapiens (human)	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.	1 (bases 1 to 1201)

**AUTHORS** Li, W.B., Gruber, C., Jesse, J. and Polayes, D.  
**TITLE** Full-length cDNA libraries and normalization  
**JOURNAL** Unpublished  
**COMMENT** On Feb 13, 2001 this sequence version replaced g1:12793308

Genoscope - Centre National de Séquençage  
Contact: [uscoscope@genoscope.cns.fr](mailto:uscoscope@genoscope.cns.fr)  
BP 191 91006 EVRY cedex - France  
Email: [segret@genoscope.cns.fr](mailto:segret@genoscope.cns.fr), Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr)  
was normalized. Library was constructed by Life Technologies, a  
division of Invitrogen. This sequence belongs to sequence cluster  
1287.f. For more information about this cluster, see  
<http://www.genoscope.cns.fr>

ncgi-bioinformatics-genoscope-csi00d006dd09qpl&cluster=1287.f. Contact :  
Feng Liang Email : [liang@lifetech.com](mailto:liang@lifetech.com) URL :  
<http://fulllength.invitrogen.com/> Invitrogen Corporation 1600  
Faraday Avenue Genoscope sequence ID : CSI00D006DD09QPL.  
Location/Qualifiers

FEATURES	SOURCE
Location/Qualifiers	
1. 1201	
/organism="Homo sapiens"	
/mol_type="mrna"	
/db_xref="taxon:9606"	
/clone="CSDD006YH18"	
/tissue_type="NEUROBLASTOMA COT 50-NORMALIZED"	
/clone_lib="Homo sapiens NEUROBLASTOMA COT 50-NORMALIZED"	
/note="1st strand cDNA was primed with a NotI-oligo (dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."	
319 a	286 c 274 g 299 t 23 others
BASE COUNT	
ORIGIN	

Alignment Scores:	
Pred. No.:	3,53e-162
Score:	1599.00
Percent Similarity:	96.65%
Best Local Similarity:	96.04%
Query Match:	88.88%
DB:	9
Length:	1201
Matches:	315
Conservative:	2
Mismatches:	11
Indels:	0
Gaps:	0

US-09-745-506-37 (1-350) x AL529615 (1-1201)

Qy	1	MethispleuYsAlaleuSerSerleuAspaphaIsSerleuSerphalaGu	20
Db	171	ATGATTTGAAAGCTCTCCCTTTCTCTGAAAGACTTTGACATCCCTCGTTTCTAG	23
Qy	21	SetrTpaSpaAnaGlyleuLeuValGluProSerProPheHisThrValAsnThrLeu	40
Db	231	AGTAGGACAAATGTGGATTACTGGHGGAAACCAAGCCCAACATACGTAAATACACTC	29
Qy	41	PheleuthrAsnAspleuthrGluGluValMetGluGluValLeuGlnLysLysAlaAsp	60
Db	291	TTCTGTGCACATGACCTCGACTGAGGAAGATGAGAGAGCTCTCGAAAGAAGCAAC	35
Qy	61	LeuileuSerThrYHisProPheArgProMetLysArgIleThrTpaAsnThr	80
Db	351	CTCATTTCTCTTCCATCCATCCGCGCTTACTTCCGACCCATGAAGCGCATACCTGGGAACACA	41
Qy	81	TryPlyGluArgleuValIleArgAlaLeuGluAsnArgValGlyLeuTyrSerProHis	10
Db	411	TGGAAAGAGGCGCTGGTGATCCGGCGCTGGAGAAACAGAGTGGATCTACCTCCCAT	47
Qy	101	ThrIleAlaTyrAspAlaAlaProGlnGluValAsnAspTyrLeuAlaLysGlyleuGlyAla	12
Db	471	ACACCCCTATATGCTGGGCCCGCCAGGCGCTCAACATCGTGTGGCTTAAGAGGCTTGAGCT	53
Qy	121	CysThrSerArgProIleHisIProSerLysAlaProAsnTyrProThrGluGlyAsnHis	14
Db	531	TGTACCTCCAGGCCCATACATCTCTCCAAAGCTCCCACTACCCCTACAGAGGAAACAC	59
Qy	141	ArgValAlaLuhAsnValAsnTyrThrGlnAspLeuAspLysValMetSerAlaValLys	16
Db	591	CGATTAACAAATCAACGTTAACCTACACCCCAAGACTGGACAAAGCTCATGTCTCCAGCTGAA	65

OY		161	GlylaespjyValSerIalThrSerpheSerIalArgThrGlyAanGlugInThr	180
Db		651	GGAAATTACCGGTGTTTCTGCACACTCTTTCTCTCTAGACTGGTAATGAGACAACA	710
OY		181	ArglleasLeuaaScyTrhgInLysAlaleuMetGlnValValaaPheLeuSerArg	200
Db		711	CGGAATTATCATCGAATTGTACTCAGAAGGCTTTGATCGAGGTGATGATTTTCCTCCCGG	770
OY		201	AsnLygGlnLeuTyrgInLysThrGlnIlleuSerLeuGluLysProLeuLeuHis	220
Db		771	AACAACAACACTTATATCAAGAACGGAATTCTGTACACGAGGAAGCCCTTGCTTCAT	830
OY		221	ThrIlyMetGlyAagLeuCyStrhleuaspGlsuSerValSerLeuAlathMetIleasp	240
Db		831	ACTGGATGGGACCGGTATATGACACACTGGATTAATCTGTCTCCCGCACCATGAMTGAT	890
OY		241	ArglleysArghIsleuLysLeuSerHisIleArgleuAlaleuGlyValGlyArgThr	260
Db		891	CGAATATAAAAACACACCTMAAACATATTCATATTCGCTTAGCCCTTGCGGTGGGAGAAC	950
OY		261	LeuGIuSerGlnValLysValValAlaleuLysAlaGlySerGlySerValLeuGln	280
Db		951	TTAAGTCTCAAGTCACAAAGTGTGGGCCCTGTGTGGCGGTTGTGGAGACGGTCTGCAG	1010
OY		281	GlyValGIuAlaaspLeuTyrlLeuThrGlyGluMetSerHisIhaSpThrhleuaspAla	300
Db		1011	GGTGTGGAGGCTGACCTTACCCTCAGGTAGAMGTCCCAMCMAGAAATTTTGATGCT	1070
OY		301	AlaseRngIlylleasNvalIlleuCygluhIsSerAenThrGluArgGlyPheLeu	320
Db		1071	GCTTCCCACARATTAATGTATCTCTGTGAAACACACACTGACGAGSGSKTYTT	1130
OY		321	SeraspLeuarGasPmetLeuasp	328
Db		1131	TCTGACCTTCGAGAAACMGCTGGAT	1154
RESULT 2				
AK011670				
LOCUS		AK011670	1376 bp	mRNA linear HTC 05-DEC-2002
DEFINITION			Mus musculus 10 days embryo whole body cDNA, RIKEN full-length enriched library, clone:2610034D10 product:nvgl interacting factor 3-like 1 (S. pombe), full insert sequence.	
ACCESSION		AK011670		
VERSION		AK011670.1	GI:12847943	
KEYWORDS			HTC; CAP trapper.	
SOURCE			Mus musculus	
ORGANISM			Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
REFERENCE				
AUTHORS			Carninci,P. and Hayashizaki,Y.	
TITLE			High-efficiency full-length cDNA cloning	
JOURNAL			Meth. Enzymol. 303, 19-44 (1999)	
MEDLINE			99279253	
PUBMED			10349636	
REFERENCE				
AUTHORS			Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.	
TITLE			Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes	
JOURNAL			Genome Res. 10 (10), 1617-1630 (2000)	
MEDLINE			20493374	
PUBMED			11042159	
REFERENCE				
AUTHORS			Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Kono,H., Akiyama,J., Nishi,K., Kitsunai,T., Teshiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,G., Sakaguchi,S., Ikegami,T., Kasaiwagi,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watabiki,M., Yoneda,Y., Isilikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.	
TITLE			RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multichipillary sequencer	

sequencing pipeline with 384 multicapillary sequencer



||||:|||||  
Db 606 GGGGTGGAGTGTCTGTCTACTCTTCTTCTCCAGGTGTGATGCTGAAGCAAC 665  
Qy 181 Arg1leasnleuasnCysThrGlnLysAlaLeuMetGlnValAlaAspPheLeuSerArg 200  
Db 666 CGGATTCAGCCGGAATTTGACTACACAGACTTTATGCAAGTGCAGAGCTTTTCCAG 725  
Qy 201 AsnLysGlnLeuYrGlnLysThrGlnLysLeuSerLeuGlnLysProLeuLeuHis 220  
Db 726 GACAGACAACTTTATCAGAAATGAAATCTTTCATTTGAAAGCCCTTCTTTGCAT 785  
Qy 221 ThrGlnMetGlyArgLeuCysThrLeuAspGlnSerValSerLeuAlaThrMetLeasp 240  
Db 786 ACTGGAATGGAGCGGTGTGTGCACACTGATGAATCTGCTCCGCAATATATAGAG 845  
Qy 241 Arg1leYsrArgH1sLeuLysLeuSerHis1leArGlnLeuAlaLeuGlyValArgThr 260  
Db 846 CGAATCAAAACACACCTTAAGCTGTGCAATCTTCCGCTTAAGCTTTGGAGTGGAGACA 905  
Qy 261 LeuGlnSerGlnValLysValAlaAlaLeuCysAlaGlySerGlySerValLeuGln 280  
Db 906 TTAGAGTCCCAAGCAAGTGTGGCCCTGTGTCTGTCTGGGGGCGAGTGTCTACAA 965  
Qy 281 GlyValGlnAlaAspLeuYrLeuThrGlnLysMetSerHisAspThrLeuAspAla 300  
Db 966 GGAATGGAGCGGACCTTACCTCAAGTGAATGTCACCAAGATGTTCTGATGCT 1025  
Qy 301 AlaSerGlnGlyLeuAsnVal1leLeuGlnLysSerAsnThrGlnArgGlyPheLeu 320  
Db 1026 GCTTCCAAAGGATCAATGTATCTCTTGTGACACACACACACTGAAGAGGCTCTT 1085  
Qy 321 SerAspLeuArgAspMetLeuAspSerHisLeuGlnAsnLys1leAsn1leLeuSer 340  
Db 1086 TCTGAGCTTCAAGAAATCTGGTCTTCATTTGAGATTAAGATTATCATTCCTGCT 1145  
Qy 341 GlnThrAspArgAspProLeuGlnValVal 350  
Db 1146 GAGACAGACAGGACCCCTCTCGTGTGTT 1175

RESULT 3  
BM545164 1117 bp mRNA linear EST 20-FEB-2002  
LOCUS AGENCOURT.6497454 NIH\_MGC\_125 Homo sapiens cDNA IMAGE:5588980  
DEFINITION AGENCOURT.6497454 NIH\_MGC\_125 Homo sapiens cDNA IMAGE:5588980  
ACCESSION BM545164  
VERSION BM545164.1 GI:18777026  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 1117)  
AUTHORS NIH-MGC http://mgi.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: csapbs-remail.nih.gov  
Tissue Procurement: Invitrogen  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: L14M12360 row: P column: 05  
High quality sequence stop: 734.

FEATURES  
Source Location/Qualifiers  
1..1117  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5588980"  
/lab\_host="DH10B"

/clone\_11b="NIH\_MGC\_125"  
/note="Organ: ovary (pool of 3); Vector: pCMV-Sport6;  
Site\_1: EcoRV (destroyed); Site\_2: NotI; RNA source pool  
of three ovaries, from females ranging in age from 38 to  
49 yo. Library is oligo-dT primed and directionally cloned  
(EcoRV site is destroyed upon cloning). Average insert  
size 2.1 kb, insert size range 1-3.5 kb. Library is  
normalized and enriched for full-length clones and was  
constructed by C. Gruber (Invitrogen). Research Genetics  
tracking code 036."

BASE COUNT 278 a 282 c 260 g 293 t 4 others  
ORIGIN  
Alignment Scores:  
Pred. No.: 9.82e-156 Length: 1117  
Score: 1539.00 Matches: 312  
Percent Similarity: 95.43% Conservative: 1  
Best Local Similarity: 95.12% Mismatches: 10  
Query Match: 85.55% Indels: 5  
Gaps: 0  
US-09-745-506-37 (1-350) x BM545164 (1-1117)

Qy 1 MetAspLeuYsrAlaLeuLeuSerSerLeuAsnAspPheAlaSerLeuSerPheAlaGln 20  
Db 88 ATGGATTTGAAGGCTCTCTCTTCTTCTTCTGTAATGACTTTGCATCTCTCTGCTGAG 147  
Qy 21 SerTrpAspAsnValGlyLeuLeuValGlnProSerProProHisThrValAsnThrLeu 40  
Db 148 AGTTGGGACATGTTGGATTACTGTTGAGAACCAAGCCACCAACATGTAATATACACTC 207  
Qy 41 PheLeuThrAsnAspLeuThrGlnGlnValMetGlnValAlaLeuGlnLysLysAlaAsp 60  
Db 208 TTCTGACCAATGACTCTGACTGAGAACTGATGAGGAGGCTGCTGCACAAAGAGCGAGAC 267  
Qy 61 Leu1leLeuSerYrHisProPro1lePheArgProMetLysArgGlyLeuThrAsnThr 80  
Db 268 CTCATCTCTCTCCATCCATCCGCTTCTTCTGACCAAGGAGGATTAACCTGGACACA 327  
Qy 81 TrpLysGlnArgLeuVal1leArgAlaLeuGlnAsnArgValGlyLeuSerProHis 100  
Db 328 TGGAGAGAGCCCTGCTGATCTCGGCTCTGAGAACACAGTCGATCTGCTCAT 387  
Qy 101 ThrAlaYrAspAlaAlaProGlnGlyLysAsnTrpLeuAlaLysGlyLeuGlyAla 120  
Db 388 ACAGCCTATGATGCTGCGCCCGCAGGCGCTCAACAACCTGTTGCTTAAAGGCTTGAGCT 447  
Qy 121 CysThrSerArgPro1leHisProSerLysAlaProAsnYrProThrGlnGlyAsnHis 140  
Db 448 TGTACCTCCAGGCCCATCATCTCTCCAAAGCTCCCACTACCTACAGAGGAAACAC 507  
Qy 141 ArgValGlnPheAsnValAsnYrThrGlnAspLeuAspLysValMetSerAlaValLys 160  
Db 508 CGAGTAGAATTCACAGTTAACTACACCCACAGACCGGACAAAGCATGTCTGACGTAA 567  
Qy 161 Gly1leAspGlyValSerValThrSerPheSerAlaArgThrGlyAsnGlnGlnThr 180  
Db 568 GGAATTTAGCGGTGTTCTGTCACTTCTTCTGTAGACTGTGTAAGAGGACCAACA 627  
Qy 181 Arg1leAsnLeuAsnCysThrGlnLysAlaLeuMetGlnValAlaAspPheLeuSerArg 200  
Db 628 CGGATTAATCTGAAATTTGACTACAGAGCTTTGATGCGAGTGTGATTTCTTCCCG 687  
Qy 201 AsnLysGlnLeuYrGlnLysThrGlnLysLeuSerLeuGlnLysProLeuLeuHis 220  
Db 688 AACCAACAACTTTATCAGAACGGAATTTGTCTGACAGGAGAGCCCTTCTCTACAT 747  
Qy 221 ThrGlnMetGlyArgLeuCysThrLeuAspGlnSerValSerLeuAlaThrMetLeasp 240  
Db 748 ACTGGAATGGAGCGGTGTGTGCACATGATGATGATCTGCTCCGCGACACATGATGAT 807  
Qy 241 Arg1leYsrArgH1sLeuLysLeuSerHis1leArGlnLeuAlaLeu-GlyValGlyArgThr 260  
Db 807

```

Db      808 CGAATANAGACACCTAAACTATCTATTCGCTTAGCCCTTTGGGGTGGGAGAAC 867
Qy      260 rleucguserginvallysvalala-leucysalalysergyservallleug 280
Db      868 CTTAAGCTCAAGCTCAAGAGTGTGGCCCTGTGTCTGTCTGTGGACAGTTCCTGC 927
Qy      280 lnglyvalgluialaspleuetyrleuthglglumetserhishisapthir-leuasr 299
Db      928 AGTGTGTCAAGCTCAAGAGTGTGGCCCTGTGTCTGTGTGGACAGTTCCTGC 987
Qy      300 Alalaaserlinglylleasn-valilleucysgluhiserasn-thrgluaraglyp 319
Db      988 GCTGCTTCCCAAGATFNAATGGCCATCCTGTGTGAACACAGCAACCTGNAACGAGCT 1047
Qy      319 heuSeraspLeu 323
Db      1048 TTCTTCTGACCTT 1061

RESULT 4
LOCUS   BX446370 1201 bp mRNA linear EST 22-MAY-2003
DEFINITION BX446370 Homo sapiens PLACENTA Homo sapiens CDNA clone CL0BA0042B11
5-PRIME, mRNA sequence.
ACCESSION BX446370
VERSION   BX446370.1 GI:31023721
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1201)
AUTHORS  Li, M.B., Gruber, C., Jesssee, J. and Polayes, D.
TITLE    Full-length cDNA libraries and normalization
JOURNAL  Unpublished
COMMENT   Contact: Genoscope
          Genoscope - Centre National de Sequencage
          BP 191 91006 Evry cedex - France
          Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
          Library was constructed by Life Technologies, a division of
          Invitrogen. This sequence belongs to sequence cluster 1287.f For
          more information about this cluster, see
          http://www.genoscope.cns.fr/
          cgi-bin/cluster.cgi?seq=CL0BA0042B11RPL&cluster=1287.f. Contact :
          Feng Liang Email : fliang@lifetech.com URL :
          http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
          Faraday Avenue Genoscope sequence ID : CL0BA0042B11RPL.

FEATURES
         source
         1..1201
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="CL0BA0042B11"
            /tissue_type="PLACENTA"
            /clone_lib="Homo sapiens PLACENTA"
            /note="Vector: PCWVSORT_6; 1st strand cDNA was primed
            with a NotI-oligo(dT) primer. Five prime end enriched,
            double-strand cDNA was digested with Not I and cloned into
            the Not I and EcoRV sites of the PCWVSORT 6 vector.
            Library was not normalized."

BASE COUNT 349 a 256 c 266 g 296 t 34.others
ORIGIN
Alignment Scores:
Pred. No.: 3 83e-155 Length: 1201
Score: 1534.00 Matches: 300
Percent Similarity: 99.348 Conservative: 2
Best Local Similarity: 98.688 Mismatches: 1
Query Match: 85.278 Indels: 1
DB: 13 Gaps: 0
US-09-745-506-37 (1-350) x BX446370 (1-1201)
Qy      48 GluGluaIaMetGluGluValleuGluLysLysAlaAspLeuIleLeuSerYrHisPro 67

```

```

Db      67 GATGAAGTATGAGAGAGTCTGCGAAAGAGGAGAGACTATTCTCTCAACATCGG 126
Qy      68 ProIlePheArgProMetLysArgIleThrPAsnThrTrpLysGluArgLeuValIle 87
Db      127 CCTATCTCCGACCCCATGGAAGCGCATACCTGGAAACATGTGAAGAGCGCTGTGATTC 186
Qy      88 ATGAlaLeuGluAsnArgValGlyIleTyrserrhishisapthirleuasrpalalapro 107
Db      187 CGGGCTGTGAGAACAGATGGGTATCTACTCTCCATACAGCTATGATGCTGGCCC 246
Qy      108 GlnGlyValAsnAsnTrpLeuAlaLysGlyLeuGlyAlaCysThrSerArgProIleHis 127
Db      247 CAGGGGCTCAACACAGTGTGGTAAAGGCTTGAGGCTTGAACCTCCAGGCCCATACAT 306
Qy      128 ProSerLysAlaProAsnTrpProThrGluGlyAsnHisArgValGluPheAsnValAsn 147
Db      307 CCTTCCAAAGCTCCCACTACACCTACAGAGGGAAACACCGAGTGAATTCACACGTTAAC 366
Qy      148 TyrThrGlnAspLeuAspLysValMetSerAlaValIleAspGlyValIleAspVal 167
Db      367 TACACCCAAAGACCTGGACAAAGTATCTGCGATGAAGAAATGACGGTGTCTGTC 426
Qy      168 ThrSerPheSerAlaArgThrGlyAsnGluGluGlnThrArgIleAsnLeuAsnCysThr 187
Db      427 ACTCTTTTCTGCTAGACTGCTGATGAGGAACAACAGGATTAACTGAATGTACT 486
Qy      188 GlnLysAlaLeuMetGlnValAlaAspPheLeuSerArgAsnLysGlnLeuTrpGlnLys 207
Db      487 CAGAAAGCTTTGTGAGGTGTGATTTCTTCCCGGAAACAAACACTTATATCAGAAG 546
Qy      208 ThrGluIleLeuSerLeuGluLysProLeuLeuHisThrGlyMetGlyArgLeuGly 227
Db      547 ACGGAATTTCTGCTACATGAGAGAGCTTTGCTTCTACATACATGGAATGGGAGGTTATGC 606
Qy      228 ThrLeuAspLysLeuValSerLeuAlaThrMetIleAspArgIleLysArgHisLeuLys 247
Db      607 ACACGTGAGTAATCTGTCTCCCGGCAACATGATGATGAATAAAGACACCTAATA 666
Qy      248 LeuSerHisIleLeuArgLeuAlaLeuGlyValGlyArgThrLeuGluSerGlnVal 267
Db      667 CTATCTCATATTCGCTTAGACCTTGGGGGGAGAACTTGAAGTCTCAAGCTCAAGAATGC 726
Qy      268 ValAlaLeuCysAlaGlySerGlySerSerValLeuGlnGlyValGluAlaAspLeuTrp 287
Db      727 GTGGCCCTGTGTCTGTCTGTGGACAGGTTCTCAGAGGTGTAGGCTGACCTTAC 786
Qy      288 LeuThrGlyGluMetSerHisHisAspThrLeuAspAlaAlaSerGlnGlyIleAsnVal 307
Db      787 CTCACAGAGTGAATGTCCTCATCATGATCTTGGATGAGTGCCTCCCAAGGAATAAATGTC 846
Qy      308 IleLeuCysGluHisSerAsnThrGluValGlyPheLeuSerSerPheLeuArgAspMetLeu 327
Db      847 ATCTCTGTGAAACAGCAACACTGAACGAGGCTTTCTTCTGACCTTGACGATATGCTG 906
Qy      328 AspSerHisLeuGluLysLysIleAsnIleIle-LeuSerGluThrAspArgAspProLe 347
Db      907 GATTTCACTTGGAGAAATAGAAATAATATATCTTATCAGAGACTGAAGAGACCTCT 966
Qy      347 uGlnValVal 350
Db      967 TCAGGTGTGTA 976

RESULT 5
LOCUS   AL581997 1201 bp mRNA linear EST 01-JUN-2003
DEFINITION AL581997 Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED
Homo sapiens CDNA clone CSDBL003Y24 5-PRIME, mRNA sequence.
ACCESSION AL581997 GI:31320228
VERSION   AL581997.2
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens

```

Eukaryote: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:  
 Mammalia: Eutheria: Primates: Catarrhini: Homiidae: Homo.  
 1 (bases 1 to 1201)  
 AUTHORS Li, W. B., Gruber, C., Jesse, J. and Polayes, D.  
 TITLE Full-length cDNA libraries and normalization  
 JOURNAL Unpublished  
 COMMENT On Feb 16, 2001 this sequence version replaced gi:12949550.  
 Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 Evry cedex - France  
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
 Library was constructed by Life Technologies, a division of  
 Invitrogen. This sequence belongs to sequence cluster 1287.f for  
 more information about this cluster, see  
 http://www.genoscope.cns.fr/  
 cgl-bln/cluster.cgi?seq=CS0DL003BD120P1&cluster=1287.f. Contact :  
 Feng Liang Email : fliang@lifetech.com URL :  
 http://fulllength.invitrogen.com/Invitrogen Corporation 1600  
 Faraday Avenue Genoscope sequence ID : CS0DL003BD120P1.  
 Location/Qualifiers

## FEATURES

source

1. 1201  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0DL003Y624"  
 /cell\_type="B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"  
 /cell\_line="RAMOS CELL LINE"  
 /clone\_id="Homo sapiens B CELLS (RAMOS CELL LINE) COT  
 25-NORMALIZED"  
 /note="1st strand cDNA was primed with a NotI-cligo(dt)  
 primer. Five prime end enriched, double-strand cDNA was  
 digested with Not I and cloned into the Not I and EcoR V  
 sites of the pCMVSPORT 6 vector. Library was normalized."  
 BASE COUNT 233 a 286 c 277 g 310 t 35 others  
 ORIGIN

## Alignment Scores:

Pred. No.: 2.09e-149 Length: 1201  
 Score: 1481.00 Matches: 307  
 Percent Similarity: 96.86% Conservative: 1  
 Best local Similarity: 96.54% Mismatches: 8  
 Query Match: 82.32% Indels: 6  
 DB: 9 Gaps: 1

US-09-745-506-37 (1-350) x AL581997 (1-1201)

QY 1 MetaspLeuTysAlaLeuLeuSerSerLeuAsnAspPheAlaSerLeuSerPheAlaGlu 20  
 |||||  
 DB 173 ATGATTTGAAGGCTCTCTCTTCTTCCATTGATGACTTGATCCCTCTGCTTGGCTGAG 232  
 QY 21 SerTTPAspaanValGlyLeuLeuValGluProSerProProHisThrValAsnThrLeu 40  
 |||||  
 DB 233 AGTTGGGACAAATGTTGGATTACTGTGTGACCAAGCCCAACACATCACTGTAATACACTC 292  
 QY 41 PheLeuThrAsnAspLeuThrGluGluValMetGluGluValLeuGluLeuTysAlaAsp 60  
 |||||  
 DB 293 TTCCGACCAATGACCTGACTGAGAGATGATGGAGGGCTGCTGCCAAGAAGGACACAC 352  
 QY 61 LeuLeuLeuSerTyrHisProProIlePheArgProMetLysArgIleThrTTPAsnThr 80  
 |||||  
 DB 353 CTCATATCTCTCTACCAATCCGCTATCTTCGACCAAGGATGAAGGCGATTAACCTGAAACA 412  
 QY 81 TrpLysGluArgLeuValIleArgAlaLeuGluAsnArgValGlyIleTyrSerProHis 100  
 |||||  
 DB 413 TGGAGAGAGCCCGCTGGTATCCGGGCTCTGAGAAACAGATCGGTATCTACTCTCTCAT 472  
 QY 101 ThrAlaTyrAspAlaAlaProGlnGlyValAsnAsnTTPLeuAlaTysGlyLeuGlyAla 120  
 |||||  
 DB 473 ACAGCTATGATGCTGCGCCCGGAGGCGCTCAACACACTGTTGGCTAAAGGCGTTGGAGCT 532  
 QY 121 CysThrSerArgProIleHisProSerLysAlaProAsnTyrProThrGluGluYasnHis 140  
 |||||  
 DB 533 TGTACCTCCAGGCCCATCATCTCTCAAGAGCTCCCAACTACCTACAGAGGAAACAC 592

QY 141 ArgValGluPheAsnValAsnTyrThrGlnAspLeuAspLysValMetSerAlaValLys 160  
 |||||  
 DB 593 CGAGTAAATTCACAGCTTAACCTACACCAAGACCTGACAAAGTCATGCTGCGATAA 652  
 QY 161 GlyLeuAspGlyValSerValThrSerPheSerAlaArgThrGlyAsnGluGluThr 180  
 |||||  
 DB 653 GGAAATTGACGGTGTCTTCTGCTACCTCTTTCTCTGCTAGACTGTTATGAGAACAAACA 712  
 QY 181 ArgLeuAsnLeuAsnCysThrGlnLysAlaLeuMetGluValValAspPheLeuSerArg 200  
 |||||  
 DB 713 CGGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 772  
 QY 201 AsnLysGluLeuTyrGlnLysThrGluLeuSerLeuGluLysProLeuLeuHis 220  
 |||||  
 DB 773 AACAAACAACTTTATTCAGAGACGGAATTCGTGATCGAGAGACCTTGTCTTACAT 832  
 QY 221 ThrGlyMetGlyArgLeuCysThrLeuAspLysSerValSerAlaThrMetLeuAsp 240  
 |||||  
 DB 833 ACTGGAATGGAGCGGTTATGTCACACATGATGATGATGATGATGATGATGATGATGAT 892  
 QY 241 ArgLeuAsnArgHisLeuLysLeuSerHisIleArgLeuAla-LeuGlyValGlyArgTh 260  
 |||||  
 DB 893 CGAATTAATAAGACACCTAAACCTATTCATATGCTTATGCTTATGCTTATGCTTATGCT 952  
 QY 260 rLeuGluSerGlnValLysValAlaAlaLeuCysAlaGlySerGlySerSerValLeuG 280  
 |||||  
 DB 953 YTTAGAGCTCTCAATCAAGTCGAGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 1012  
 QY 280 nGlyValGluAlaAspLeuTyrLeuThrGlyLysMetSerHisAspThrLeuAspAl 300  
 |||||  
 DB 1013 GGGTGTGAGGCTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTAC 1070  
 QY 300 aAlSerGlnGlyLeuAsnValIleLeuCysGluHisSerAsnThrGluArg 317  
 |||||  
 DB 1071 -GCYTCACCAAGCATTA-AATGTM---WCCTCTGTGAMAAACAACTAGAACGA 1117

## RESULT 6

AK003978

LOCUS

DEFINITION

Mus musculus 18-day embryo whole body cDNA, RIKEN full-length

enriched library, clone:110030G24 product:1991 interacting factor

3-like 1 (S. pombe), full insert sequence.

ACCESSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

141 ArgValGluPheAsnValAsnTyrThrGlnAspLeuAspLysValMetSerAlaValLys 160  
 |||||  
 593 CGAGTAAATTCACAGCTTAACCTACACCAAGACCTGACAAAGTCATGCTGCGATAA 652  
 161 GlyLeuAspGlyValSerValThrSerPheSerAlaArgThrGlyAsnGluGluThr 180  
 |||||  
 653 GGAAATTGACGGTGTCTTCTGCTACCTCTTTCTCTGCTAGACTGTTATGAGAACAAACA 712  
 181 ArgLeuAsnLeuAsnCysThrGlnLysAlaLeuMetGluValValAspPheLeuSerArg 200  
 |||||  
 713 CGGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 772  
 201 AsnLysGluLeuTyrGlnLysThrGluLeuSerLeuGluLysProLeuLeuHis 220  
 |||||  
 773 AACAAACAACTTTATTCAGAGACGGAATTCGTGATCGAGAGACCTTGTCTTACAT 832  
 221 ThrGlyMetGlyArgLeuCysThrLeuAspLysSerValSerAlaThrMetLeuAsp 240  
 |||||  
 833 ACTGGAATGGAGCGGTTATGTCACACATGATGATGATGATGATGATGATGATGATGAT 892  
 241 ArgLeuAsnArgHisLeuLysLeuSerHisIleArgLeuAla-LeuGlyValGlyArgTh 260  
 |||||  
 893 CGAATTAATAAGACACCTAAACCTATTCATATGCTTATGCTTATGCTTATGCTTATGCT 952  
 260 rLeuGluSerGlnValLysValAlaAlaLeuCysAlaGlySerGlySerSerValLeuG 280  
 |||||  
 953 YTTAGAGCTCTCAATCAAGTCGAGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 1012  
 280 nGlyValGluAlaAspLeuTyrLeuThrGlyLysMetSerHisAspThrLeuAspAl 300  
 |||||  
 1013 GGGTGTGAGGCTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTAC 1070  
 300 aAlSerGlnGlyLeuAsnValIleLeuCysGluHisSerAsnThrGluArg 317  
 |||||  
 1071 -GCYTCACCAAGCATTA-AATGTM---WCCTCTGTGAMAAACAACTAGAACGA 1117

AK003978 1467 bp mRNA linear HTC 05-DEC-2002  
 Mus musculus 18-day embryo whole body cDNA, RIKEN full-length  
 enriched library, clone:110030G24 product:1991 interacting factor  
 3-like 1 (S. pombe), full insert sequence.  
 AK003978  
 AK003978.1 GI:12834961  
 HTC; CAP trapper.  
 Mus musculus (house mouse)  
 Mus musculus  
 Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 Carninci, P. and Hayashizaki, Y.  
 High-efficiency full-length cDNA cloning  
 Meth. Enzymol. 303, 19-44 (1999)  
 99279253  
 10349636  
 2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
 Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
 Normalization and subtraction of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new genes  
 Genome Res. 10 (10), 1617-1630 (2000)  
 20499374  
 11042159  
 3  
 Shibata, K., Itoh, M., Aizawa, K., Nagao, S., Sasaki, N., Carninci, P.,  
 Kono, H., Akiyama, J., Nishi, K., Katsunuma, T., Tashiro, H., Itoh, M.,  
 Suni, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,  
 Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,  
 Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matabiki, M.,  
 Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,  
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
 RIKEN integrated sequence analysis (RISA) system--384-format





OY		180	PATJ1ELASNUISNCYSPTHRIUNLYSALALEMEGLINVALVLAASPHEUSERR	200
Db		756	CCGGATCAGCGCTGAATTGTACTCAGAAGACCTTTATGACGGTGTTAACCTTTTCGCCA	815
OY		200	GASNLSGLNLEUTYRGUINYSTRHGUILLLEUSERLEUGLULSPROLEULEUH1	220
Db		816	GGACAGCAACAATTATGCAGAAACGTGAATAATCTTTCAATTGGAGAACCTTGCTTCGCA	875
OY		220	STHG1WMCGLYRGLEUCYSTPHRLIEUAAPGLUSERVALSERILEUMIARHTMETILIAS	240
Db		876	TACGGAAATGGACGGTGTGTGCACACAGCATGAATCTGTCTCCCTGGCAATAAATGATAGA	935
OY		240	PARG1IELYSARGHISLEUYLSLEUSERHIS1EARGDEUALALEUNG1VALGIARYTH	260
Db		936	CGCAATCAAACACACCCTAAAGCTGTGCGACATCTTCGCTTAGCTCTGGAGTGGGAGAAC	995
OY		260	RLEUGLSERGINVALYSVALVAIALEUCYALAGLYSERGISSESRVALLEUG1	280
Db		996	ATTAGAGCCCCAAGTCAAAAGTTGGGCCCTGTGTGGCTTGCTGGGGCACTGTTCTPACA	1055
OY		280	NGLYVALGLUALAASPLEUTYRLTEUTHRG1GLIMETESERHISHISAPTRHLEUASP1	300
Db		1056	AGGATGGAGGCCGACCTTTCACGTACACAGGTGAATGTCCACACATGATGTTCGATGTC	1115
OY		300	AALSERG1NCLYLIESNVALLILEUCYSGLUHISERSNTHRIGUARGLYPHELE	320
Db		1116	TGCTYTCCAAGGATGCAATGTCAATCTTTTGGAACACAGCAACACTGAAGAGGCTTCCT	1175
OY		320	USERASPLeutrygaspmetleuaaspserrhisleuglunsnlysliaenilleileuse	340
Db		1176	TTCTGAGCTTAAAGAAATGCTGGGTGTCTCACCTTGAGATATAGATTACATTATCTCTCTC	1235
OY		340	rGlutThraspargasproleuginvalva1	350
Db		1236	TGAGACAGACAGGACCTCTCCGTGTGTGT	1266
RESULT 7				
LOCUS	BX393871	930 bp	mRNA	linear EST 13-MAY-2003
DEFINITION	BX393871 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens			
ACCESSION	CDNA clone CS0DC005YM21 5'-PRIME, mRNA sequence.			
VERSION	BX393871			
KEYWORDS	BX393871.1 GI:30624084			
SOURCE	EST.			
ORGANISM	Homo sapiens (human)			
REFERENCE	Homo sapiens			
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
TITLE	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
JOURNAL	L1.W.B., Gruber.C., Jesssee.J. and Polayves.D.			
COMMENT	Full-length cDNA libraries and normalization			
	Unpublished			
	Contact: Genoscope			
	Genoscope - Centre National de Sequencage			
	BP 191 J1006 EVRY cedex - France			
	Email: seqref@genoscope.cns.fr Web : www.genoscope.cns.fr			
	Library was constructed by life technologies, a division of			
	Invitrogen. This sequence belongs to sequence cluster 1287.f for			
	more information about this cluster, see			
	http://www.genoscope.cns.fr/			
	cgl-bin/cluster.cgi?seq-CS0DC005AG11Q1&cluster=1287.f. Contact :			
	Feng liang Email : fliang@lifetech.com URL :			
	http://lifetech.invitrogen.com/ Invitrogen Corporation 1600			
	Paradise Avenue Genoscope sequence ID : CS0DC005AG11Q1.			
FEATURES				
source	Location/Qualifiers			
	1..930			

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CSDDC005YX21"
/tissue_type="NEUROBLASTOMA COT 25-NORMALIZED"
/clone_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"
(note="1st strand cDNA was primed with a NotI-oligo(3T)

```

BASE COUNT	248 a	231 c	214 g	237 t
ORIGIN	primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."			

Alignment Scores:	
Pred. No.:	1.97e-138
Score:	1378.00
Percent Similarity:	100.00%
Best Local Similarity:	100.00%
Query Match:	76.60%
DB:	13
Length:	933
Matches:	26
Conservative:	0
Mismatches:	0
Indels:	0
Gaps:	0

US-09-745-506-37 (1-350) x BX393871 (1-930)

QY	MetAspLeuPheValAlaLeuSerSerLeuAsnAspPheAlaSerLeuSerPheAlaGlu	20
Db	130 ATGATATTGAAGGCTCTCCCTTTCTTCTTGAATGACTTTGGATCCCTCTGTTGCTGAG	18
QY	21 SerTrpAspAsnValGlyLeuLeuValGluProSerProProHisThrValAsnThrLeu	40
Db	190 AGTTGGACAATGTTGGATTACTGGTGGAAACCAAGCCACACATACATGTAATATACATC	24
QY	41 PheLeuThrAsnAspLeuThrGluGluValMetGluGluValLeuGlnIleValAlaAsp	60
Db	250 TTCTGTGCACATACCTGTGACTGAGAAAGATGAGAGAGGTCTCTCAAAAGAACGACAC	30
QY	61 LeuIleLeuSerThrHisProProIlePheAspGProMetLysArgIleThrTrpAsnThr	80
Db	310 CTGATTTCCTCCATACATCGCGCTACTCTCCGACCCATGAAGCGCATACCTGGAAACCA	36
QY	81 TrpLysGluArgLeuValIleArgAlaLeuGluAsnValGlyIleLysSerProHis	100
Db	370 TGGAAAGAGCGCCCTGGTGATCCGGCGCTCGGAGAACAGATGGTATCTACTCTTCAT	42
QY	101 ThrAlaTrpAspAlaAlaProGlnGlyValAsnAsnTrpLeuAlaLysGlyLeuGlyVala	120
Db	430 ACAGCCATATATGCTGCGGCCCGCCAGGGCGTCMACACTGGTTGGCTTAAAGGCGTTGAGCT	48
QY	121 CysThrSerArgProIleHisProSerLysAlaProAsnTyrProThrGluGlyAsnHis	140
Db	490 TGTACCTCCAGGGCCCATACATCTTCCAAAGCTCCCACTACCTTCACAGAGGAACAC	54
QY	141 ArgValGluPheAsnValAsnTyrTrpGlnAspLeuAspLysValMetSerAlaValLys	160
Db	550 CGAGTAAATTCACACGTTTACCTACACCCACAGACTGACCAAGTATGCTGTCACTGAA	60
QY	161 GlyIleAspGlyValSerValThrSerPheSerAlaArgThrGlyAsnGluGlnThr	180
Db	610 GGAATTGACGGTGTCTGTCTCACTTCTTTTCTGTCAGACGCTGAATGAGAACAAACA	66
QY	181 ArgIleAsnLeuAsnGlySerGlnLysAlaMetMetGlnValValAspPheLeuSerArg	200
Db	670 CGGATTAATCTGAATTTACTCAAGAAAGGCTTGATGATCAGGTGATGATTTTCTCCGG	72
QY	201 AsnLysGlnLeuTyrGlnLysTrpGluIleLeuSerLeuGlnLysProLeuLeuHis	220
Db	730 AACCAAAACACTTATATCAGAGACGGAATTCGTCTACTGAGAGAGCCTTGCTTCTACAT	78
QY	221 ThrGlyMetGlyArgLeuCysThrLeuAspGluSerValSerLeuAlaThrMetLeasp	240
Db	790 ACTGGAATGGACGCGTATATCACACATGAGATCTGTCTCCCTGGCAACCATGATTTAT	84
QY	241 ArgIleLysArgHisLeuLysLeuSerHisIleArgLeuAlaLeuGlyValGlyArgThr	260
Db	850 CGAATTAAGAACACCTTAACATATCTCATATTCGCTTAGCCCTTGGGGTGGGGAAC	90
QY	261 LeuGluSerGlnValLysVal 267	
Db	910 TTACAGTCTCAGTCAAGTCAAGTC 930	



BUI72435  
 LOCUS BUI72435 857 bp mRNA linear EST 04-SEP-2002  
 DEFINITION AGENCOURT 7970438 NIH\_MGC\_72 Homo sapiens cDNA clone IMAGE:6164871  
 5', mRNA sequence.  
 ACCESSION BUI72435  
 VERSION BUI72435.1 GI:22686419  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 857)  
 NIH-MGC <http://mgc.nci.nih.gov/>.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
 Tissue Procurement: ATCC/DCPD/BTP  
 cDNA Library Preparation: Life Technologies, Inc.  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: L1AM13522 row: k column: 16  
 High quality sequence stop: 728.  
 Location/Qualifiers  
 1..857  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:6164871"  
 /tissue\_type="melanotic melanoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH\_MGC-72"  
 /note="Organ: Skin; Vector: pCMV-SPORT6; Site\_1: NotI;  
 Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dr.  
 Average insert size 2 kb. Library constructed by Life  
 Technologies."  
 BASE COUNT 234 a 201 c 199 g 223 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 7.26e-137 Length: 857  
 Score: 1363.00 Matches: 267  
 Percent Similarity: 100.008 Conservative: 0  
 Best Local Similarity: 100.008 Mismatches: 0  
 Query Match: 75.768 Indels: 0  
 Gaps: 0  
 US-09-745-506-37 (1-350) x BUI72435 (1-857)  
 QY 84 ArgLeuValIleArgAlaLeuGluAsnArgValGlyIleTyrSerProHsrhAlaTyr 103  
 DB 3 CGCCTGGATATCCGGGCTCTGGAGAACAGATCGTATCTCTCTCTACAGCCTAT 62  
 QY 104 AspAlaAlaProGlnGlyValAsnAsnTPIleuAlaAlaGlyGlyLeuGlyAlaGlySer 123  
 DB 63 GATGCTGCCGCCCGAGGGCGTCAACAACATGCTGGCTAAAGGGCTTGAGCTGTGACCTCC 122  
 QY 124 ArgProIleHisProSerLysAlaProAsnTyrProThrGlnGlyAsnHisArgValGlu 143  
 DB 123 AGGCCCATATCATCTCCAAAGCTCCCACTACCCACAGAGAGAAACACAGAGTAA 182  
 QY 144 PheAsnValAsnTyrThrGlnAspLeuAspLysValMetSerAlaValLysGlyIleAsp 163  
 DB 183 TTCACAGCTTAACATACCCAGACCTCGAGCAAAAGTCATCTGCACTGAAGGAATTGAC 242  
 QY 164 GlyValSerValThrSerPheSerAlaArgThrGlyAsnGlnGlnGlnGlnGlnGlnGln 183  
 DB 243 GGTGTTTCTGTCACTTCTTTCTGCTAGAGCTGGTAATGAGAACAAACGAGATTAAAT 302  
 QY 184 LeuAsnCysThrGlnLysAlaLeuMetGlnValValAspPheLeuSerArgAsnLysGln 203

DB 303 CTGAATTGTATCTCAGAAAGCTTTGATGACAGGTGTGATTTCTTCCCGGAAACAACA 362  
 QY 204 LeuTyrGlnLysThrGlnIleLeuSerLeuGlnLysProLeuLeuHisThrGlyMet 223  
 DB 363 CTTTATCAGAAAGAGGAAATCTGTCTACCTGAGAACGCTTGTCTTACTAGGAATG 422  
 QY 224 GlyArgLeuCysThrLeuAspGluSerValSerLeuAlaThrMetIleAspArgIleLys 243  
 DB 423 GGAGGTTATGACACCTGGATGATCTGTCTCCCTGGCAACATGATGATGCATTAATA 482  
 QY 244 ArgHisLeuLysLeuSerHisIleArgLeuAlaLeuGlyValGlyArgThrLeuGluSer 263  
 DB 483 AGACACTAAACTATCTCATATTCCTTATGCCCTTGGGGTGGGAGAACTTAGAGTCT 542  
 QY 264 GlnValLysValValAlaLeuCysAlaGlySerGlySerSerValLeuGlnGlyVal 283  
 DB 543 CAAGTCAAAATCCGTGGCCCTGTGTGCTGGTGTGGAGCAGCGCTTCTGCAGGGTGTGAG 602  
 QY 284 AlaAspLeuTyrLeuThrGlnGlyMetSerHisAspThrLeuAspAlaAlaSerGln 303  
 DB 603 GCTGACCTTTAACCCTCAGAGTGAATGTCCATCATGATCTTGGATGCTGCTCCAA 662  
 QY 304 GlyIleAsnValIleLeuCysGlnHisSerAsnThrGlnArgGlyPheLeuSerAspLeu 323  
 DB 663 GGAATTAATGTCTATCTCTGTGAAACACAGCAACACATGACGAGCTTCTTGTGACCTT 722  
 QY 324 ArgAspMetLeuAspSerHisLeuGluAsnLysIleAsnIleLeuSerGlnThrAsp 343  
 DB 723 CGAATATAGCTGTGATCTCTGCTGAGAAATGATTAATATCTATCAGAGACTGAC 782  
 QY 344 ArgAspProLeuGlnValVal 350  
 DB 783 AGGAGCCCTTCTCAGGTGTA 803  
 RESULT 9  
 BUI859307 929 bp mRNA linear EST 16-OCT-2002  
 LOCUS BUI859307  
 DEFINITION AGENCOURT\_10422791 NIH\_MGC\_109 Homo sapiens cDNA clone  
 IMAGE:6650148 5', mRNA sequence.  
 ACCESSION BUI859307  
 VERSION BUI859307.1 GI:24044299  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 929)  
 NIH-MGC <http://mgc.nci.nih.gov/>.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
 Tissue Procurement: ATCC  
 cDNA Library Preparation: Rubin Laboratory  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: L1CM2894 row: o column: 12  
 High quality sequence stop: 754.  
 Location/Qualifiers  
 1..929  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:6650148"  
 /tissue\_type="teratocarcinoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH\_MGC-109"  
 /note="Organ: Ovary; Vector: pORF7; Site\_1: EcoRI; Site\_2:  
 XhoI; cDNA made by oligo-dr priming. Directionally cloned



|||||  
Db 325 CTCATCTCTCTACCATCGGCTATCTTCCGACCCATGAAGCCATACCTGGAAACA 384  
Oy 81 TTPlySGluArgLeuValIleArgAlaLeuGluAsnArgValGlyIleTyrSerProHis 100  
Db 385 TGGAGAGAGCGCTGGTGGATCCGGCTCTGGAGAACAGACTCGGTATCTCTCCCTCAT 444  
Oy 101 ThAlaTyrAspAlaAlaProGlnGlyValAsnAspTyrLeuAlaIleGlyLeuGlyAla 120  
Db 445 ACAGCCTATGATCTGCGCCCGCCAGGGCGTCAACACAGGCTTGGCTAAAGGGCTGGAGCT 504  
Oy 121 CysThrSerArgProIleHisProSerTyrAlaProAsnTyrProThrGlnGlyAsnHis 140  
Db 505 TGTACCTCCAGGCCCATACATCTCTCCAAAGCTCCCAACATACCTACAGAGGAAACAC 564  
Oy 141 ArgValGluPheAsnValAsnTyrThrGlnAspLeuAspIleValMetSerAlaValLys 160  
Db 565 CGAGTGAATTCACCTTAACTTACCTACCCCAAGACCTGGACAAAGTCAATGTCAGTGAAA 624  
Oy 161 GlyIleAspGlyValSerValThrSerPheSerAlaArgThrGlnGlyAsnGlnGlnThr 180  
Db 625 GGAATTGACGGCTTTCTCTGCTACTTCTTTTCTGCTAGGACTGGTGGTGGAGAACAAACA 684  
Oy 181 ArgIleAsnLeuAsnCysThrGlnLysAlaLeuMetGlnValValAspPheLeuSerArg 200  
Db 685 CGGATTAATCTGAATGTACTCAGAAAGCTTGTATGAGAGTGGATTTTCTTCCCGG 744  
Oy 201 AsnIleGlnLeuTyrGlnLysThrGlnIleLeuSerLeuGlnLysProLeuLeuLeuHis 220  
Db 745 AACAAACAACTTATATCAGAAAGAGGAATTCGTCACTGAGAAAGCTTGTGCTTCATCAT 804  
Oy 221 ThrGlyMetGlyArgLeuCysThrLeuAspGlySerValSerLeuAlaThrMetIleAsp 240  
Db 805 ACAGGATGGAGCGGTATGACACAGTGAATGATCTCTCCCTGGCAACACATGATTTGAT 864  
Oy 241 ArgIleLysArgHisLeuLysLeuSerHisIleArgLeuAlaLeuGlnValGlyArgThr 260  
Db 865 CGAATTAAGAAAGACCACTAANAACTATCTCATATTCGCTTACCTTGGGGGGGGAGAAC 924  
Oy 261 -leuGlnSerGlnValLysValAlaIle-leuCysAlaGlySerGlySer 277  
Db 925 CTTAGAGTCTCAAGTCAAGTCCGTGGCCCTGTGTGCTGGTGGAGAAC 977

RESULT 11  
BQ229243 836 bp mRNA linear EST 02-MAY-2002  
LOCUS AGENCOURT\_7510797 NIH\_MGC\_72 Homo sapiens cDNA clone IMAGE:6055296  
DEFINITION 5', mRNA sequence.  
ACCESSION BQ229243  
VERSION BQ229243.1 GI:20410643  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 836)  
AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgaps@remail.nih.gov](mailto:cgaps@remail.nih.gov)  
Tissue Procurement: ATCC/DC/DMP  
CDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.lnl.gov>  
Plate: L1AM1315 row: f column: 01  
High quality sequence stop: 677.  
Location/Qualifiers  
1..836  
source /organism="Homo sapiens"

/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6055296"  
/issue\_type="melanotic melanoma"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_id="NIH\_MGC\_72"  
/note="Organ: skin; Vector: pCMV-Sport6; Site: 1; Notif.  
Site: 2; Sali: Cloned unidirectionally. primer: Oligo dT.  
Average insert size 2 kb. Library constructed by Life  
Technologies."  
BASE COUNT 218 a 205 c 200 g 211 t 2 others  
ORIGIN  
Alignment Scores:  
Pred. No.: 1,3e-136 Length: 836  
Score: 1360.50 Matches: 268  
Percent Similarity: 97.45% Conservative: 0  
Best Local Similarity: 97.45% Mismatches: 5  
Query Match: 75.63% Indels: 2  
DB: 13 Gaps: 1  
US-09-745-506-37 (1-350) x BQ229243 (1-836)  
Oy 3 LeuLysAlaLeuLeuSerSerLeuAsnAspPheAlaSerLeuSerPheAlaGlySerTrp 22  
Db 2 TTGAGGCTCTCTCTTCTTCTTCTTGAATGACATTGTCATCCCTCTGCTTGGTGGAGCTGG 61  
Oy 23 AspAsnValGlyLeuLeuValGluProSerProHisThrValAsnThrLeuPheLeu 42  
Db 62 GACAAATGTTGGATTACTGCTGGAGAACCAAGCCACACATCACTGTAATTAACATCTTCTG 121  
Oy 43 ThrAsnAspLeuThrGlnGlnGlnValMetGlnGlnValLeuGlnLysLysAlaAspLeuIle 62  
Db 122 ACCAATACCTGACTGAGAGATGATGAGAGAGTGTGCAAAAGAGCGACACTGATT 181  
Oy 63 LeuSerTyrHisProProIlePheArgProMetLysArgIleThrPAsnThrTrpLys 82  
Db 182 CTCTCCATCACCCTCGGCTATCTTCCGACCATGAAAGGCATTAACCTGGAACATGAGAG 241  
Oy 83 GluArgLeuValIleArgAlaLeuGluAsnArgValGlyIleTyrSerProHisThrAla 102  
Db 242 GAGCGCTGTGTATCCGGCTCTGGAGAACAGACTCGGTATCTCACTCTCCATACAGACC 301  
Oy 103 TyrAspAlaAlaProGlnGlyValAsnAspTyrLeuAlaIleGlyLeuGlnValAlaCysThr 122  
Db 302 TATGATCTCTGCGCCCGCCAGGCGTCAACACATGCTTGAAGGCTTGGAGCTTGTACC 361  
Oy 123 SerArgProIleHisProSerTyrAlaProAsnTyrProThrGlnGlyAsnHisArgVal 142  
Db 362 TCCAGGCCCATACATCTCTCCAAAGCTCCCAACTACCTTCAGAGGAAACACACGAGTA 421  
Oy 143 GluPheAsnValAsnTyrThrGlnAspLeuAspIleValMetSerAlaValLysGlyIle 162  
Db 422 GAATTCACAGCTTAACCTACCCCAAGACCTGACAAACTCATGTCGTGAGTGAAGGAAATT 481  
Oy 163 AspGlyValSerValThrSerPheSerAlaArgThrGlnGlnGlnGlnGlnThrArgIle 182  
Db 482 GAGCGTCTTCTGCTACTTCTTTTCTGCTAGAGCTGTAATGAGAACAAACCGGATT 541  
Oy 183 AsnLeuAsnCysThrGlnLysAlaLeuMetGlnValValAspPheLeuSerArgAsnLys 202  
Db 542 AATCTGAATGTACTCAGAAAGCTTGTATGATGAGTGTGATATTTCTTCCCGGAACAAA 601  
Oy 203 GlnLeuTyrGlnLysThrGlnIleLeuSerLeuGlnLysProLeuLeuHisThrGly 222  
Db 602 CAATTTATCAGAAAGAGGAATTCGTCACTGAGAGAGCTTGTGCTTCACTATCAGGA 661  
Oy 223 MetGlyArgLeuCysThrLeuAspGlySerValSerLeuAlaThrMetIleAspArgIle 242  
Db 662 ATGGAGAGGTATGACACATGATGATGATGCTCTCCCTGGCAACCATGATGATGATA 721  
Oy 243 LysArgHisLeuLysLeuSerHisIleArgLeuAlaLeuGlnValGlyArgThrLeuGln 262  
|||||



Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr  
was normalized. Library was constructed by Life Technologies, a  
division of Invitrogen. This sequence belongs to sequence cluster  
1287.f For more information about this cluster, see  
http://www.genoscope.cns.fr/

cgi-bin/cluster.cgi?seq=CSDD006DD09NP1&cluster=1287.f. Contact :  
Feng Liang Email: fliang@lifetech.com URL :  
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
Faraday Avenue Genoscope sequence ID: CSDD006DD09NP1.

## FEATURES

source

Location/Qualifiers

1..1201

/organism="Homo sapiens"

/mol\_type="mrna"

/db\_xref="taxon:9606"

/clone="CSDD006YH18"

/issue\_type="NEUROBLASTOMA COT 50-NORMALIZED"

/clone\_lib="Homo sapiens NEUROBLASTOMA COT 50-NORMALIZED"

/note="1st strand cDNA was primed with a NotI-oligo(dT)  
primer. Five prime end enriched, double-strand cDNA was  
digested with Not I and cloned into the Not I and EcoR V  
sites of the pCMVSPORT 6 vector. Library was normalized."

## BASE COUNT

325 a 268 c 278 g 298 t 32 others

## ORIGIN

## Alignment Scores:

Pred. No.:	1..34e-133	Length:	1201
Score:	1335.00	Matches:	277
Percent Similarity:	87.65%	Conservative:	7
Best Local Similarity:	85.49%	Mismatches:	37
Query Match:	74.21%	Indels:	4
DB:	9	Gaps:	0

US-09-745-506-37 (1-350) x AL563496 (1-1201)

QY 30 GluProSerProthiThrValAsnThrLeuPheLeuThrAsn-AspLeuThrGlu1 49  
 DB 1194 GAACCAAGCCGCCACAMCCCTGTAATAAATTTCTCTACACATGATGCGCCCTGAGGA 1135  
 QY 49 uValMetGluGluValLeuGlnIleuLysAlaAspLeuLeuSerThrHisProPoi1 69  
 DB 1134 AGTGATGGGGGGGGG-CTGCAAAAAGAGAGAGCCCTMATTTTCTCTCCATCCGCTTT 1076  
 QY 69 ePhe-ArgProMetLysArgGlieThrTrpAsnThrTrpLysGluArgLeuValIleArg1 89  
 DB 1075 CTTCACACCCCAATAGAGCGCATACCTGGAACACATGAAAGGGCGCTGTATCCGGG 1016  
 QY 89 lAlaLeuLysAsnArgValGlyIleLeuSerProHisThrAlaTrpAspAlaAlaProGln1 109  
 DB 1015 CTCTGGAGAACAGAGTCGGTTTCTCTCTCATACAGCTTGTATSTGCCCCCAGG 956  
 QY 109 lYValAsnAsnTrpLeuAlaLysGlyLeuGlyAlaCysThrSerArg-ProIleHisPro 128  
 DB 955 GCGTCAACAACCTGTTGGCTAAAGGGCTTGTCCTGTCAGGCCCATACACATCC 896  
 QY 129 SerLysAlaProAsnTrpProThrGluGlyAsnHisArgValAluPheAsnValAsnTrp 148  
 DB 895 TCCAAAGCCGCCCACTACCTACAGAGGGAACACACGATTAATTAATCAAGTTTACTAC 836  
 QY 149 ThrGlnAspLeuAspLysValMetSerAlaValIleGlyLysPheGlyValSerValThr 168  
 DB 835 ACCCAAGACCTGGACACRAGTCATKCTCTGCAAGTAAGGAATTCAGCGTGTCTGCAC 776  
 QY 169 SerPheSerAlaArgThrGlyAsnGluGluGlnThrArgTleAsnLeuAsnGlyThrGln 188  
 DB 775 TCTTTAAAAAATGAAGCTGTATATGAGAACACACACGATTAATTCATGATCTACTAG 716  
 QY 189 LysAlaLeuMetGlnValValAspPheLeuSerArgAsnLysGlnLeuTrpGlnLysThr 208  
 DB 715 AAGGCTTTGATGAGAGTGTATTTCTTCCCGAACACACACTTTTTCAGAGAGAG 656  
 QY 209 GluIleLeuSerLeuGlnLysProLeuLeuLeuHisThrGlyMetGlyArgLeuCysThr 228  
 DB 655 GAAATTCGTCACTGAGAAAGCCCTTTCATCAACATACGTGAATGGAGCGTTATGACACA 596

QY 229 LeuAspGluSerValSerLeuAlaThrMetIleAspArgGlieLysArgHisLeuLysLeu 248  
 DB 595 CTGAGATGATCTGTCTCCCGGCAACCATATGATGAAATTAAGAACACCTTAACACTC 536  
 QY 249 SerHisIleArgLeuAlaLeuGlyValGlyArgThrLeuGluSerGlnValIleValVal 268  
 DB 535 TCCCATATTCGCTTACGCCCTTGGGGTGGGAGAACCTTAAGTCTCAAGCAAGTGTG 476  
 QY 269 AlaLeuCysAlaGlySerGlySerValLeuGlnGlyValGluAlaAspLeuTrpLeu 288  
 DB 475 GCCCTGTCTCTGTTCTCTGGGGCAGCGCTTCAGAGGCTTGAAGCTGACCTTCCCTC 416  
 QY 289 ThrGlyGluMetSerHisAspThrLeuAspAlaAlaSerGlnGlyIleAsnValIle 308  
 DB 415 ACAGGTGAGTGTCTCTCATGATTTTGTGATGCTCTTCCCAAGAAATTAATGTCATC 356  
 QY 309 LeuCysGluHisSerAsnThrGluArgGlyPheLeuSerAspLeuArgAspMetLeuAsp 328  
 DB 355 CTCTGTGAACACGCGCCACTGAACGAGGCTTCTTCTGACCTTTCGAGTTATGCTGAT 296  
 QY 329 SerHisIleuLysAsnLysIleAsnIleIleLeuSerGluThrAspArgAspProLeuGln 348  
 DB 295 TCTCATTGGAGATTAAGATTAATATTATCTTTCAGAGACTAAGACGACCTCTTCAG 236  
 QY 349 ValVal 350  
 DB 235 GTGCTA 230  
 RESULT 14  
 AK076788  
 LOCUS  
 DEFINITION  
 Mus musculus adult male testis cDNA, RIKEN full-length enriched  
 library, clone:4930455H01 product:99g1 Interacting factor 3-like 1  
 (S. Pombe), full insert sequence.  
 AK076788  
 VERSION  
 AK076788.1 GI:26345659  
 KEYWORDS  
 HTC; CAP trapper.  
 SOURCE  
 Mus musculus (house mouse)  
 ORGANISM  
 Mus musculus  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1  
 Carninci, P. and Hayashizaki, Y.  
 High-efficiency full-length cDNA cloning  
 Meth. Enzymol. 303, 19-44 (1999)  
 99279253  
 JOURNAL  
 MEDLINE  
 PUBMED  
 10349636  
 2  
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
 Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.  
 Normalization and subtraction of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new genes  
 Genome Res. 10 (10), 1617-1630 (2000)  
 2049374  
 11042159  
 3  
 Shibata, K., Itoh, M., Aizawa, K., Nagaoke, S., Sasaki, N., Carninci, P.,  
 Konno, H., Akiyama, J., Nishi, K., Kusunagi, T., Tashiro, H., Itoh, M.,  
 Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishimura, T., Harada, A.,  
 Yamamoto, R., Matsumoto, H., Sakaue, S., Ikegami, T., Kashiwagi, K.,  
 Fujisawa, S., Inoue, K., Togawa, Y., Iizawa, M., Ohara, E., Watabiki, M.,  
 Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,  
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.  
 RIKEN Integrated sequence analysis (RISA) system--384-format  
 sequencing pipeline with 384 multicapillary sequencer  
 Genome Res. 10 (11), 1757-1771 (2000)  
 20530913  
 JOURNAL  
 MEDLINE  
 PUBMED  
 11076861  
 4  
 Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,  
 Aikawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S.,  
 Aizawa, K., Iizawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamana, I.,

	SAITO,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R., Kido,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T., Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochwa,H., Kuehl,P., Lewis,S., Matsuo,Y., Nikiado,I., Pesole,G., Quackenbush,J., Schriml,L.M., Stabili,F., Suzuki,R., Tomita,M., Wagner,L., Washio,T., Sakai,K., Stabili,F., Suzuki,R., Tomita,M., Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N., Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C., Fletcher,C., Fujita,A., Gariboldi,M., Gustincich,S., Hill,D., Hofmann,M., Hume,D.A., Kamlay,M., Lee,N.H., Lyons,P., Marchionni,L., Mashima,J., Mazzarelli,D., Mommaerts,P., Norgone,P., Riing,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H., Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H., Toyohara-Oka,K., Wang,K.H., Weitz,C., Whitaker,C., Wilmberg,L., Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohzuki,S. and Hayashizaki,Y.
TITLE	Functional annotation of a full-length mouse cDNA collection
JOURNAL	Nature 409 (6821), 685-690 (2001)
MEDLINE	21085660
PUBMED	11217851
REFERENCE	5
AUTHORS	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL	Nature 420, 563-573 (2002)
REFERENCE	6 (bases 1 to 1919)
AUTHORS	Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A., Arawaka,T., Bono,H., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hirooka,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Kono,H., Konda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ono,M., Okazaki,Y., Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shibata,Y., Shihagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toyohara,T., Yamamura,T., Yamana,I., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE	Direct Submission
JOURNAL	Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama Kanagawa 230-0045, Japan (E-mail:genome-resgsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT	cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Please visit our web site for further details. URL:http://genome.gsc.riken.go.jp/ URL:http://fantom.gsc.riken.go.jp/.
FEATURES	Location/Qualifiers
SOURCE	1. 1919
	/organism="Mus musculus"
	/mol_type="mRNA"
	/strain="C57BL/6J"
	/db_xref="FANTOM,DB:493045SH01"
	/db_xref="MGI:1908024"
	/db_xref="taxon:10090"
	/clone="493045SH01"
	/sex="male"
	/tissue_type="testis"
	/clone_lib="RIKEN full-length enriched mouse cDNA library
	/dev_stage="adult"
	144. 1166
CDS	/note="unamed protein product; Ng91 interacting factor 3-like 1 (S. pombe) (MGDI:MGI:1929485, GBI NM_022988, evidence: BLASTN, 99%, match=1373) putative

	polyA_signal	polyA_site	BASE COUNT	ORIGIN	Alignment Scores:	Pred. No.:	Score:	Percent Similarity:	Best Local Similarity:	Query Match:	DB:	US-09-745-506-37 (1-350) x AK076788 (1-19129)																																																																																
QY	1	MetaspLeuTysAlaLeuLeuSerSerLeuAsnAspPheAlaSerLeuSerPheAlaGlu	20	222	ATGATCTGAAGAGCGCTCTCTCCGCTTGAATGACTTTCGATCCCTCATTTCTGCTG	281	21	SerTrpAspAsnValGlyLeuLeuValGluProSerProProHisThrValAsnThrLeu	40	282	AGCTGGGACAAATGGGGAATCTGGTGGAGCCCAAGCCCAACCCCAATGTAATACATCTC	341	41	PheLeuThrAsnAspLeuThrGluGluValMetGluGluValLeuGluGlnLysLysAlaAsp	60	342	TTCCGACCAATGACCTGACGAGGAGGTGATGACAGACAGCGCTGCAAAAGAGGAGAC	401	QY	61	LeuLeuLeuSerTyrHisProProLeuPheArgProMetLysArgIleThrTrpAsnThr	80	402	TTTCATTTCTCTCTACCATCCACCTCTTTTCCGCCCATGACAGACATTAATCTGGAAAC	461	QY	81	TrpLysGluArgLeuValIleArgAlaLeuGluAsnArgValGlyIleTyrSerProHis	100	462	TGGAGGAGGTCTCTGTAATCCGGGCTCTGGGAACAGAGTGGCTGTACTCTCTCCAC	521	Db	101	ThrAlaTyrAspAlaAlaProGlnGlyValAsnAsnTrpLeuAlaLysGlyLeuGluAla	120	QY	121	CysThrSerArgProIleHisProSerLysAlaProAsnTyrProThrGluGluAsnHis	140	Db	522	ACAGCCTATGATGCGCGCACCCAGGAGAGTCAACACTGGTGGCCAAAGGCGTGGAACT	581	QY	141	ArgValGluPheAsnValAsnTyrThrGlnAspLeuAspLysValMetSerAlaValLys	160	Db	582	TGCATACACGCGCCATCCACCTTCCAGAGCTCCAGATACCACCAAGAGGAGGCTCAC	641	QY	161	GlyIleAspGlyValSerValThrSerPheSerAlaArgTyrGlyAsnGluGluGlnThr	180	Db	642	CGACTAGATTTAGTGTGAACCGCACGACCAAGACTGGAGCAAAAGTCAATGCTACACTGAGA	701	QY	181	ArgIleAsnLeuAsnCysThrGlnLysAlaLeuMetGlnValValAspPheLeuSerArg	200	Db	702	GCGGTGGAGGTCTCTCTCTACTCTTTTCCGCCAGGTGATGATGTGAAGGCAAAAC	761	QY	201	AsnLysGlnLeuTyrGlnLysThrGluIleLeuSerLeuGluLysProLeuLeuLeuHis	220	Db	762	CGGATACAGCCTTAATGTACTCAGAAAGACTTTTGATGCAGGTGCTGACTTCTTCTCCAG	821	QY	221	ThrGlyMetGlyArgLeuCysThrLeuAspGluSerValSerLeuAlaThrMetIleAsp	240	Db	822	GACACACAACTTATACAGAAACTGAAATCTTTCATTGGAAGAAAGCTTGTGCTTCGAT	881	QY	241	ThrGlyMetGlyArgLeuCysThrLeuAspGluSerValSerLeuAlaThrMetIleAsp	260	Db	882	GACACACAACTTATACAGAAACTGAAATCTTTCATTGGAAGAAAGCTTGTGCTTCGAT	941

|||||  
Db 882 ACTGGATGGAGCGGTGTGCACACTGATGATCTCTCCCTGGCAATGAAGAGAG 941  
Qy 241 ArgilleysArgHisLeuLysLeuSerHisIleArgLeuAlaLeuGlyAlaGlyArgThr 260  
Db 942 CGAATCAAAACACACTTAAGCTGTGCATCTTCGCTTACCTCTTGAGTGGGAGAACCA 1001  
Qy 261 LeuGlusSerGlnValLysValValAlaLeuCysAlaGlySerGlySerSerValLeuGln 280  
Db 1002 TTAGAGTCCCAAGTCAAAAGTTGGCCCTGTGTCTGTCTGGGGGAGAGTTCTACAA 1061  
Qy 281 GlnValGlnAlaSerLeuThrGlyGlnMetSerHisHisAspThrLeuAsp 299  
Db 1062 GGAGTGGAGCGGACCTTACCTACACAGCAGTGTCTGTGAACAGAGACGTGAA 1118  
RESULT 15  
AL521920 957 bp mRNA linear EST 22-MAY-2003  
LOCUS AL521920 Homo sapiens NEUROBLASTOMA COT 10-NORMALIZED Homo sapiens  
DEFINITION CDNA clone CS0DB003YPL4 5-PRIME, mRNA sequence.  
ACCESSION AL521920  
VERSION AL521920.2 GI:31040194  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 957)  
L1,W.B., Gruber,C., Jesse,J. and Polayes,D.  
Full-length cDNA libraries and normalization  
Unpublished  
On Feb 13, 2001 this sequence version replaced gi:12785413.  
COMMENT  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 Evry cedex - France  
Email: sequef@genoscope.cns.fr Web : www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 1287.f For  
more information about this cluster, see  
http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CS0DB003DH07OP1&cluster=1287.f Contact :  
Feng Liang Email : fliang@lifetech.com URL :  
http://fulllength.invitrogen.com/ invitrogen Corporation 1600  
Paradise Avenue Genoscope sequence ID : CS0DB003DH07CPL  
FEATURES  
source  
1..957  
Location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DB003YPL4"  
/issue\_type="NEUROBLASTOMA COT 10-NORMALIZED"  
/clone\_lib="Homo sapiens NEUROBLASTOMA COT 10-NORMALIZED"  
/note="1st strand cDNA was primed with a NotI-oligo(dT)  
primer. Five prime end enriched, double-strand cDNA was  
digested with Not I and cloned into the Not I and EcoR V  
sites of the pCMVSPORT 6 vector. Library was normalized."  
BASE COUNT 250 a 239 c 217 g 250 t 1 others  
ORIGIN  
Alignment Scores:  
Pred. No.: 1,36e-131 Length: 957  
Score: 1315.00 Matches: 261  
Percent Similarity: 98.86 Conservative: 0  
Best Local Similarity: 98.86 Mismatches: 3  
Query Match: 73.10 Indels: 2  
DB: 9 Gaps: 0  
US-09-745-506-37 (1-350) x AL521920 (1-957)  
Qy 1 MetAspLeuLysAlaLeuLeuSerLeuAsnAspPheAlaSerLeuSerPheAlaGlu 20  
Db 166 ATGATTTGAAGGCTCTCTTCTTCTTCTTGAATGACTTTGCAATCCCTCTGCTTGTCTGAG 225

Qy 21 SerTPAspAsnValGlyLeuLeuValGlnProSerProHisThrValAsnThrLeu 40  
Db 226 AGTTGGACAAATGTGGATTACTGTGGAAACCAAGCCACACATGTAATACACTC 285  
Qy 41 PheLeuThrAsnAspLeuThrGlnGlnValMetGlnGlnValLeuGlnLysValAsp 60  
Db 286 TTCCTGACCAATGACTGACTGAGAGATGATGAGAGGCTGCAAAAGAAAGGAGAC 345  
Qy 61 LeuIleLeuSerThrHisProProIlePheArgProMetLysArgIleThrPAsnThr 80  
Db 346 CTCATCTCTCCACACATCCGCGCCCTATCTCCAGACCATGAAGGCAATACCTGGAAACA 405  
Qy 81 TPPLysGlnArgLeuValIleArgAlaLeuGlnAsnArgValGlyIleThrSerProHis 100  
Db 406 TGGAGAGAGCGCTGGATCCGGGCTCTGGAACAGAGTCCGTAATCTACTCTCCAT 465  
Qy 101 ThrAlaThrAspAlaIleProGlnGlyValAsnAspThrPheValLysGlyLeuGlyAla 120  
Db 466 ACAGCTATGATCTCGCGCCCGCCAGGCGCTCAACACTGTTGGCTAAAGGCTTGGAGCT 525  
Qy 121 CysThrSerArgProIleHisProSerLysAlaProAsnThrProThrGlnGlyAsnHis 140  
Db 526 TGTACCTCCAGGCCCATACATCTCTCCAAAGCTCCCACTACCTACAGAGGAAACAC 585  
Qy 141 ArgValGlnPheAsnValAsnThrThrGlnAspLeuAspLysValMetSerAlaValLys 160  
Db 586 CGAGTAGAATTCACAGCTTAACACCCAGACCTGGACAAAGTCAATGCTGCAGTGA 645  
Qy 161 GlyIleAspGlyValSerValThrSerPheSerAlaArgThrGlnGlnGlnThr 180  
Db 646 GGAATGACCGGTCTTCTGCTACTCTTTTCTGTAGAGACTGTAATGAGAAACA 705  
Qy 181 ArgIleAsnLeuAsnLysThrGlnLysAlaLeuMetGlnValValAspPheLeuSerArg 200  
Db 706 CGGATTAATGTGAATGTACTAGAAAGCTTTGATGAGCTGATGATTTCTTCCCGG 765  
Qy 201 AsnLysGlnLeuThrGlnLysThrGlnLeuSerLeuGlnLysProLeuLeuHis 220  
Db 766 AACCAACAACTTTATCAGAAAGACGGAATCTGTCACTGGAGAAAGCTTTGCTTCAAT 825  
Qy 221 ThrGlyMetGlyArgLeuLysThrIleAspGlnSerValSerLeuAlaThrMetIleAsp 240  
Db 826 ACTGGATGGAGCGTTATGACACATGATGAATCTGTCTCCCTGGCAACCATGATGAT 885  
Qy 241 ArgIleLysArgHisLeuLysLeuSerHisIleArgLeuAlaLeuGlyValGlyArgThr 260  
Db 886 CGAAT-AAAAGACACCTAAACATATCTCATATTCGCTTACCCCTGGGGGAGAGACT 944  
Qy 261 LeuGlusSerGln 264  
Db 945 TA-GAGTCTCAA 955

Search completed: August 23, 2003, 19:55:01  
Job time : 1911 secs



**THIS PAGE BLANK (USPTO)**